



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2014

---

**Texte schreiben 2013 – Standardisierte Erfassung von Schreibkompetenzen.  
Bericht zuhanden der Bildungsdirektion Kanton Zürich**

Oostlander, Jeannette ; Wespi, Barbara

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-124572>

Scientific Publication in Electronic Form

Published Version

Originally published at:

Oostlander, Jeannette; Wespi, Barbara (2014). Texte schreiben 2013 – Standardisierte Erfassung von Schreibkompetenzen. Bericht zuhanden der Bildungsdirektion Kanton Zürich. Zürich: Institut für Bildungsevaluation.



**Universität  
Zürich<sup>UZH</sup>**

**Institut für Bildungsevaluation  
Assoziiertes Institut der Universität Zürich**

---

## **Texte schreiben 2013 – Standardisierte Erfassung von Schreibkompetenzen**

Bericht zuhanden der Bildungsdirektion Kanton Zürich

Jeannette Oostlander & Barbara Wespi

Zürich, 25. April 2014

Institut für Bildungsevaluation  
Assoziiertes Institut der Universität Zürich  
Wilfriedstrasse 15  
8032 Zürich

Tel: 043 268 39 60  
Fax: 043 268 39 67  
[www.ibe.uzh.ch](http://www.ibe.uzh.ch)

[Jeannette.Oostlander@ibe.uzh.ch](mailto:Jeannette.Oostlander@ibe.uzh.ch)  
[Barbara.Wespi@ibe.uzh.ch](mailto:Barbara.Wespi@ibe.uzh.ch)

# Inhalt

<b>1</b>	<b>Ausgangslage</b>	<b>4</b>
<b>2</b>	<b>Durchführung</b>	<b>4</b>
2.1	Schreibauftrag	5
2.2	Themen	5
<b>3</b>	<b>Beurteilung der Texte</b>	<b>7</b>
3.1	Beurteilungsraster	7
3.2	Beurteilungskriterien	8
<b>4</b>	<b>Qualität des Beurteilungsrasters</b>	<b>10</b>
4.1	Analyse der Beurteilungskriterien	10
4.2	Eindimensionalität des Beurteilungsrasters	13
<b>5</b>	<b>Beurteilungszuverlässigkeit</b>	<b>13</b>
5.1	Zusammensetzung des Korrekturteams	13
5.2	Verfahren der Mehrfachkorrekturen	14
5.3	Übereinstimmung der Bewertung	14
5.4	Strenge der Bewertungen	15
5.5	Auswahl und Schwierigkeit der Themen	16
<b>6</b>	<b>Testergebnisse</b>	<b>18</b>
6.1	Verteilung der Ergebnisse auf der Stellwerkskala	18
6.2	Verteilung der Fähigkeiten der Schülerinnen und Schüler nach Geschlecht	19
6.3	Verteilung der Fähigkeiten der Schülerinnen und Schüler nach Schulstufe	19
<b>7</b>	<b>Beispieltexte zu den einzelnen Kompetenzniveaus</b>	<b>20</b>
<b>8</b>	<b>Fazit</b>	<b>29</b>
<b>9</b>	<b>Glossar: Statistische Kennwerte</b>	<b>30</b>

# 1 Ausgangslage

Seit dem Schuljahr 2010/11 wird das computergestützte Testsystem «Stellwerk» im Kanton Zürich in der 8. Klasse flächendeckend eingesetzt. Stellwerk umfasst Tests für die Fachbereiche Deutsch, Englisch, Französisch, Mathematik sowie Natur und Technik. Die Stellwerk-Tests werden ausschliesslich am Computer gelöst, wobei es sich um adaptive Tests handelt, welche sich an die Fähigkeiten der Schülerinnen und Schüler anpassen. Adaptive Tests haben den Vorteil gegenüber dem traditionellen «Papier-Bleistift-Test», dass die Objektivität der Durchführung gesichert ist und dass der Computer bei der Korrektur keine Fehler macht. Computergestützte Tests haben allerdings insofern Nachteile, als dass sie nur reproduktive Fähigkeiten testen. Produktive Fähigkeiten wie ausführliche Antworten auf offene Fragen oder ganze Texte können am Computer bis jetzt nicht zufriedenstellend korrigiert und bewertet werden.

Im Dezember 2010 hat der Bildungsrat des Kantons Zürich beschlossen, das Testsystem Stellwerk mit der Erfassung von produktiven Fähigkeiten im Fachbereich Deutsch zu ergänzen. Anhand eines standardisierten Verfahrens soll die Schreibkompetenz der Schülerinnen und Schüler beurteilt werden. Das Institut für Bildungsevaluation hat das Modul «Textproduktion Deutsch» im Kanton Zürich in den Jahren 2007 bis 2010 als Teil des Pilotprojekts «Neugestaltung 3. Sek» an insgesamt 4'192 Schülerinnen und Schülern erprobt<sup>1</sup>. Im Schuljahr 2013/14 wurde das Modul «Texte schreiben» bereits zum zweiten Mal bei allen 8. Klassen regulär durchgeführt.

# 2 Durchführung

Die Durchführung des Moduls «Texte schreiben» fand am 25. und 26. November 2013 statt. Die Teilnahme war für alle Schülerinnen und Schüler der 8. Klasse im Kanton Zürich obligatorisch. Den Schülerinnen und Schülern wurden zwei Themen vorgelegt, wovon sie ein Thema auswählen konnten. Dazu schrieben sie während 60 Minuten einen Text, der in der Regel eine bis drei Seiten umfasste. Insgesamt verfassten 10'410 Schülerinnen und Schüler einen Text. Die Texte wurden von den Lehrpersonen kopiert und die Originale dem Institut für Bildungsevaluation zur Beurteilung zugestellt.

Die Texte der Schülerinnen und Schüler wurden anschliessend am Institut für Bildungsevaluation von sieben eigens dafür geschulten Personen – Lehrpersonen, Germanistinnen und Germanisten – anhand eines standardisierten Beurteilungsrasters bewertet (eine detaillierte Darstellung des Beurteilungsrasters befindet sich im Praxisteil dieses Berichtes). Die erreichte Punktzahl im Modul «Texte schreiben» wurde in die Ergebnisse von «Stellwerk» integriert, welche von den Lehrpersonen am Computer eingesehen und ausgedruckt werden können. Zugleich wurden die beurteilten Texte an die Lehrpersonen und ihre Schülerinnen und Schüler zurückgesandt. Das ausgefüllte Be-

---

<sup>1</sup> Moser, U. (2007). Standardisierte Erfassung der sprachlichen Kompetenzen im Fachbereich «Texte schreiben». Kurzbericht zuhanden des Pilotprojekts «Neugestaltung des 9. Schuljahrs» und der Projektleitung Sekundarstufe I der Bildungsdirektion des Kantons Zürich.

Moser, U. (2008). Standardisierte Erfassung der sprachlichen Kompetenzen im Fachbereich «Texte schreiben». Kurzbericht zuhanden des Pilotversuchs «Neugestaltung des 9. Schuljahrs» und der Projektleitung Sekundarstufe I der Bildungsdirektion des Kantons Zürich.

Moser, U. & Keller, F. (2009). Standardisierte Erfassung der sprachlichen Kompetenzen im Fachbereich «Texte schreiben». Kurzbericht zuhanden des Pilotprojekts «Neugestaltung 3. Sek» und der Projektleitung der Bildungsdirektion des Kantons Zürich.

Moser, U. (2010). Standardisierte Erfassung der sprachlichen Kompetenzen im Fachbereich «Texte schreiben». Kurzbericht zuhanden des Pilotprojekts «Neugestaltung 3. Sek» und der Projektleitung der Bildungsdirektion des Kantons Zürich.

urteilungsraster wurde an jeden Aufsatz angeheftet, sodass die Beurteilung der Texte für die Lehrpersonen sowie für die Schülerinnen und Schüler einsehbar ist.

## 2.1 Schreibauftrag

Die Ausgangslage für die Entwicklung der Schreibaufträge ist der Lehrplan des Kantons Zürich. Darauf basierend wurde ein Schreibauftrag entwickelt, welcher die Richtziele zum Schreiben möglichst umfänglich abdeckt. Allerdings eignet sich nicht jeder Schreibauftrag gleich gut für eine standardisierte Erfassung der Schreibkompetenzen von Schülerinnen und Schülern. Je offener ein Auftrag formuliert wird, desto schwieriger gestaltet sich die standardisierte Beurteilung. Aus diesem Grund werden Schreibaufträge bevorzugt, welche einen Lebens- beziehungsweise Alltagsbezug haben und mit klaren Aufgaben verbunden sind.

Mit dem Modul «Texte schreiben» wird angestrebt, die Fähigkeit zu erfassen, Texte korrekt und verständlich zu formulieren, je nach Zielsetzung adressatengerecht, logisch und präzise zu formulieren sowie Sprache ästhetisch ansprechend und kreativ einzusetzen (Harsch, Neumann, Lehmann & Schröder, 2007)<sup>2</sup>.

## 2.2 Themen

Im Sinne einer standardisierten schriftlichen Anleitung wurden die Schülerinnen und Schüler aufgefordert, zu einem Sachverhalt Stellung zu nehmen. Die Themen wurden den Schülerinnen und Schülern mittels einer kurzen Einleitung und drei zu beantwortenden Fragen vorgestellt. Die Fragen dienen den Schülerinnen und Schülern als Leitfaden und helfen bei der Strukturierung der Texte. Mit der ersten Frage wurden die Schülerinnen und Schüler dazu aufgefordert, eigene Beobachtungen und Erfahrungen zum Thema zu formulieren und damit die Leserinnen und Leser ins Thema einzuführen. Zwei weitere Fragen zielten auf eine argumentative Stellungnahme zum Thema ab. Bei beiden Themen wurden die Schülerinnen und Schüler aufgefordert, einen Brief mit korrekter Anrede- und Grussformel zu verfassen.

Zur Wahl standen zwei Themen:

- freier Zugang zum Internet
- Schulreisen und Klassenlager

Die Anleitung für beide Themen des Schreibauftrags ist in den Abbildungen 1 und 2 dargestellt.

---

<sup>2</sup> Harsch, C., Neumann, A., Lehmann, R. & Schröder, K. (2007). Schreibfähigkeit. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung* (S. 42–62). Weinheim: Beltz

Abbildung 1: Schreibauftrag «Internet»

### 1. Internet

*Einige Eltern sind der Meinung, dass der freie Zugang zum Internet gefährlich sei, süchtig mache und die Jugendlichen vom Lernen ablenke. Deshalb verbieten gewisse Eltern den freien Zugang zum Internet.*

Verfasse als Schülervertreterin oder Schülervertreter einen Brief an die Eltern deiner Schule. In diesem Brief kämpfst du für den freien Zugang zum Internet.

- Welche Erfahrungen und Beobachtungen machst du zu diesem Thema?
- Weshalb soll der freie Zugang zum Internet erlaubt sein?
- Suche Lösungen: Wie soll man mit dem Internet umgehen?
  - Achte darauf, dass dein Text dem Stil eines öffentlichen Briefes entspricht.
  - Achte auf eine korrekte Anrede und Grussformel.

Abbildung 2: Schreibauftrag «Schulreise»

### 2. Schulreisen

*Die Schulleitung deiner Schule ist der Meinung: «Schulreisen und Klassenlager kosten zu viel Geld und sind zu gefährlich. Zudem geht wertvolle Lernzeit in der Schule verloren. Deshalb sollen in Zukunft keine Reisen und Lager mehr innerhalb der Schulzeit organisiert werden.»*

Verfasse als Schülervertreterin oder Schülervertreter einen Brief an die Schulleitung. In diesem Brief kämpfst du dafür, dass in der Schulzeit weiterhin Ausflüge und Lager stattfinden werden.

- Welche Erfahrungen und Beobachtungen machst du zu diesem Thema?
- Weshalb sollen Reisen und Lager während der Schulzeit auch in Zukunft stattfinden?
- Suche Lösungen: Wie sollen Schulreisen und Klassenlager in Zukunft organisiert werden?
  - Achte darauf, dass dein Text dem Stil eines öffentlichen Briefes entspricht.
  - Achte auf eine korrekte Anrede und Grussformel.

Ebenfalls zum Schreibauftrag gehören Angaben darüber, wie die Texte bewertet werden (siehe Punkt 4 in Abbildung 3). Die ersten beiden inhaltlichen Kriterien beziehen sich auf den expliziten Schreibauftrag. Zum einen muss der kommunikative Auftrag erfüllt werden, zum anderen sollte der Text nach expliziten Vorgaben strukturiert werden («Hat dein Text eine Einleitung, einen Hauptteil und einen Schluss?»). Darüber hinaus wird den Schülerinnen und Schülern auch mitgeteilt, dass die Kriterien Sprachrichtigkeit («Ist dein Text sprachlich richtig?») und Sprachangemessenheit («Ist die Wortwahl passend und abwechslungsreich?») beurteilt werden.

Abbildung 3: Anleitung zum Schreibauftrag

1. Lies die Informationen zu den zwei Themen genau durch.
2. Wähle eines der zwei Themen aus und verfasse dazu einen Text.
3. Gehe wie folgt vor:
  - Du darfst Notizpapier zur Vorbereitung gebrauchen.
  - Du darfst zuerst einen Entwurf oder ein Mindmap auf das Notizpapier schreiben.
  - Schreibe deinen Text gut lesbar auf die ausgeteilten Blätter.
  - Du darfst den Duden oder das Wörterbuch benutzen.
  - Du hast 60 Minuten Zeit.
4. Dein Text wird nach folgenden Kriterien bewertet:
  - Werden in deinem Text alle drei Fragen ausführlich beantwortet?
  - Hat dein Text eine Einleitung, einen Hauptteil und einen Schluss?
  - Ist dein Text klar und verständlich?
  - Ist dein Text sprachlich richtig?
  - Ist die Wortwahl passend und abwechslungsreich?

### 3 Beurteilung der Texte

#### 3.1 Beurteilungsraster

Zur Beurteilung der Texte im Rahmen des Moduls «Texte schreiben» wurde ein Beurteilungsraster basierend auf den vier Basisdimensionen «Inhalt», «Textaufbau», «Sprachrichtigkeit» und «Sprachangemessenheit, Schreibstil und Kreativität» entwickelt. Diese vier Dimensionen entstanden in Anlehnung an das «Zürcher Textanalyseraster» von Nussbaumer und Sieber (1994)<sup>3</sup> sowie basierend auf dem Vorgehen von Becker-Mrotzek und Böttcher (2011)<sup>4</sup>. Das Zürcher Textanalyseraster (Nussbaumer & Sieber 1994) erfasst die «sprachsystematische und orthographische Richtigkeit», die «funktionale Angemessenheit», die «ästhetische Angemessenheit» und die «inhaltliche Relevanz». Für die pädagogische Praxis erwies sich das Zürcher Analyseraster als zu detailliert. Zudem bietet es für die inhaltlichen Belange zu wenig Anhaltspunkte für die Beurteilung von Schülertexten (vgl. Neumann, 2007)<sup>5</sup>. Aus diesen Gründen schlagen Becker-Mrotzek und Böttcher (2011) einen Basiskatalog mit zwölf Kriterien vor, verteilt auf die fünf Basisdimensionen «Sprachrichtigkeit», «Sprachangemessenheit», «Inhalt», «Aufbau» und «Schreibprozess». Unter der Basisdimension «Schreibprozess» werden nach Becker-Mrotzek und Böttcher (2011) die folgenden beiden Kriterien zusammengefasst: «Planen/Überarbeiten» (Lässt der Text Planungs- und Überar-

<sup>3</sup> Nussbaumer, M. & Sieber, P. (1994). Texte analysieren mit dem Zürcher Textanalyseraster. In P. Sieber (Hrsg.), *Sprachfähigkeiten – besser als ihr Ruf und nötiger denn je! Ergebnisse aus einem Forschungsprojekt* (S. 141-186). Aarau: Sauerländer.

<sup>4</sup> Becker-Mrotzek, M. & Böttcher, I. (2011). *Schreibkompetenzen entwickeln und beurteilen*. Berlin: Cornelsen

<sup>5</sup> Neumann, A. (2007). *Briefe schreiben in Klasse 9 und 11*. Beurteilungskriterien, Messungen, Textstrukturen und Schülerleistungen. Münster: Waxmann.



beutungsspuren erkennen?) und «Wagnis/Kreativität» (Lässt der Text ein besonderes sprachliches Wagnis erkennen? Ist er in besonderer Weise kreativ?).

Anpassungen wurden entsprechend bisheriger Erfahrungen mit der Korrektur von Texten und aufgrund testtheoretischer Gütekriterien vorgenommen. Zudem wurden die inhaltlichen Kriterien auf die konkrete Aufgabenstellung angepasst. Das Beurteilungsverfahren entspricht einem analytischen Vorgehen, bei dem verschiedene Aspekte eines Textes unabhängig voneinander nach verbal formulierten Abstufungen bewertet werden (Analytical Scoring; Weigle, 2002)<sup>6</sup>.

### 3.2 Beurteilungskriterien

In Tabelle 1 sind die Kriterien zur Beurteilung der Texte bezüglich der Dimension *Inhalt: Auftragserfüllung und Aussagekraft* ersichtlich. In dieser Dimension wurden bei beiden Themen inhaltlich analoge Bereiche beurteilt. Bewertet wurde, ob die Ausführungen zum gewählten Thema passen, inwiefern auf die gestellten Fragen eingegangen wird, wie gehaltvoll die Aussagekraft des Textes ist, ob der Adressatenbezug gelingt und ob die Schülerinnen und Schüler beim Verfassen des Textes auch ein inhaltliches Wagnis eingehen.

Tabelle 1: Kriterien des Beurteilungsrasters zur Dimension «Inhalt: Auftragserfüllung und Aussagekraft»

Kriterien	Abstufungen
1.1 Ausführungen	1 passen nur teilweise zum gestellten Thema 2 passen weitgehend zum gestellten Thema 3 passen gut zum gestellten Thema
1.2 Fragen	1 werden nur teilweise beantwortet 2 werden weitgehend beantwortet 3 werden ausführlich, detailliert beantwortet
1.3 Aussagekraft / Gehalt	1 nur teilweise gehaltvoll 2 weitgehend gehaltvoll 3 gehaltvoll
1.4 Adressatenbezug	1 nur teilweise gelungen 2 weitgehend gelungen 3 gut gelungen
1.5 inhaltliches Wagnis / Kreativität	1 wagt etwas, etwas kreativ 2 wagt recht viel, recht kreativ 3 wagt viel, kreativ 4 wagt sehr viel, sehr kreativ

In Tabelle 2 sind die Kriterien zur Dimension *Textaufbau und Textzusammenhang* dargestellt. Es wurde beurteilt, ob ein Text in eine sinnvolle äussere Gliederung (Einleitung, Hauptteil, Schluss) eingeteilt ist. Zusätzlich wurde bewertet, ob der Text in sich logisch zusammenhängt (Kohärenz) und ob die Sätze und Abschnitte sprachlich sinnvoll verbunden sind (Kohäsion).

<sup>6</sup> Weigle, S. C. (2002). *Assessing Writing*. Cambridge: Cambridge University Press.

Tabelle 2: Kriterien des Beurteilungsrasters zur Dimension «Textaufbau und Textzusammenhang»

Kriterien	Abstufungen
2.1 Textaufbau (Abschnitte – äussere Gliederung)	1 zufällig, unüberlegt, ungegliedert oder inkonsequent 2 gegliedert (Einleitung, Hauptteil und Schluss sind ersichtlich)
2.2 logischer Zusammenhang (innere Gliederung – Kohärenz)	1 nur teilweise logisch ausgeführte Gedanken 2 meistens logisch ausgeführte Gedanken 3 logisch ausgeführte Gedanken
2.3 sprachlicher Zusammenhang (innere Gliederung – Kohäsion)	1 nur teilweise sinnvoll verbunden 2 meistens sinnvoll verbunden 3 sinnvoll verbunden, auch bei komplexeren Verbindungen

Tabelle 3 umfasst die Kriterien zur Dimension *Sprachrichtigkeit*. Dabei wurde beurteilt, ob ein Text in Bezug auf Rechtschreibung, Grammatik, Satzbau und Satzzeichen korrekt ist.

Tabelle 3: Kriterien des Beurteilungsrasters zur Dimension «Sprachrichtigkeit»

Kriterien	Abstufungen
3.1 Rechtschreibung	1 nur ansatzweise beherrscht 2 teilweise beherrscht 3 weitgehend beherrscht 4 nahezu fehlerfrei
3.2 Grammatik (Genus, Kasus, Tempus, Modus)	1 nur ansatzweise beherrscht 2 teilweise beherrscht 3 weitgehend beherrscht 4 nahezu fehlerfrei
3.3 Satzbau (Korrektheit, Vollständigkeit, inkl. Kohäsionsmittel)	1 nur ansatzweise beherrscht 2 teilweise beherrscht 3 weitgehend beherrscht 4 nahezu fehlerfrei
3.4 Satzzeichen (Satzschlusszeichen, Kommas, Anführungszeichen etc.)	1 nur ansatzweise beherrscht 2 teilweise beherrscht 3 weitgehend beherrscht 4 nahezu fehlerfrei

Tabelle 4 beinhaltet die Kriterien der Dimension *Sprachangemessenheit*. Bei dieser Dimension wurde bewertet, ob die Wortwahl angemessen und treffsicher ist und ob der Satzbau abwechslungsreich ist. Zudem wurde beurteilt, wie gewandt sich die Schülerinnen und Schüler ausdrücken (Sprachstil) sowie ob sie sprachlich etwas wagen.

Tabelle 4: Kriterien des Beurteilungsrasters zur Dimension «Sprachangemessenheit, Schreibstil und Kreativität»

Kriterien	Abstufungen
4.1 Satzbau	1 etwas abwechslungsreich 2 abwechslungsreich 3 sehr abwechslungsreich, vielseitig
4.2 Wortwahl	1 eher treffend, angemessen 2 treffend, angemessen 3 treffsicher, angemessen, auch bei komplexeren Begriffen
4.3 Schreibstil	1 sprachlich eher unsicher, wenig gewandt 2 sprachlich recht sicher, etwas gewandt 3 sprachlich sicher, gewandt 4 sprachlich sehr sicher, sehr gewandt
4.4 Sprachliches Wagnis – Kreativität und Ästhetik	1 wagt etwas, etwas kreativ 2 wagt recht viel, recht kreativ 3 wagt viel, kreativ 4 wagt sehr viel, sehr kreativ

## 4 Qualität des Beurteilungsrasters

### 4.1 Analyse der Beurteilungskriterien

In Tabelle 5 ist die Qualität der Beurteilungskriterien der Dimension «Inhalt» dargestellt. Die Trennschärfen<sup>7</sup> der fünf Beurteilungskriterien liegen alle über 0.30 und können somit als ausreichend bis gut bewertet werden. Die Kriterien 1.1 und 1.4 sind eher einfach ausgefallen, das bedeutet, dass sie von vielen Schülerinnen und Schülern gut erfüllt wurden. 73 Prozent der Schülerinnen und Schüler haben in ihren Texten Ausführungen gemacht, welche gut zum gestellten Thema passen (Kriterium 1.1). Der Adressatenbezug ist auch einem Grossteil der Schülerinnen und Schüler (72%) gut gelungen (Kriterium 1.4). Am schwersten fiel den Schülerinnen und Schüler das inhaltliche Wagnis beziehungsweise die inhaltliche Kreativität (Kriterium 1.5).

<sup>7</sup> Die statistischen Begriffe in diesem Kapitel werden im Glossar im Anhang dieses Berichtes erläutert.

Tabelle 5: Qualität der Kriterien zur Dimension «Inhalt»

Beurteilungskriterien	Abstufungen	Schwierigkeit	Trennschärfe	Infit (MNSQ)	T-Wert
1.1 Ausführungen	1	2%	0.52	1.14	7.3
	2	25%			
	3	73%			
1.2 Fragen	1	17%	0.57	1.17	10.6
	2	55%			
	3	28%			
1.3 Aussagekraft / Gehalt	1	10%	0.64	1.01	0.8
	2	61%			
	3	29%			
1.4 Adressatenbezug	1	5%	0.46	1.47	20.6
	2	23%			
	3	72%			
1.5 inhaltliches Wagnis / Kreativität	1	8%	0.54	1.22	12.7
	2	58%			
	3	32%			
	4	1%			

Tabelle 6 beinhaltet die Darstellung der Qualität der Beurteilungskriterien der Dimension «Textaufbau und Textzusammenhang». Die Trennschärfe aller drei Kriterien ist als zufriedenstellend bis gut zu bewerten. Die beiden Kriterien zum logischen und sprachlichen Zusammenhang passen gut zum Raschmodell, während das Kriterium zum Textaufbau mit einem erhöhten Infit Wert gerade noch in einem akzeptablen Bereich liegt. 58% der Schülerinnen und Schüler gliederten ihren Text so, dass Einleitung, Hauptteil und Schluss ersichtlich waren. Bei den beiden Kriterien zum logischen und sprachlichen Zusammenhang verteilen sich die Texte relativ gleichmässig auf die beiden oberen Abstufungen (2 und 3), während die unterste Stufe (1) seltener vertreten ist. Dies bedeutet, dass nur wenige Texte aus nur teilweise logischen Gedanken bestanden und die Sätze und Abschnitte eher selten nur teilweise sinnvoll verbunden waren.

Tabelle 6: Qualität der Kriterien zur Dimension «Textaufbau und Textzusammenhang»

Beurteilungskriterien	Abstufungen	Schwierigkeit	Trennschärfe	Infit (MNSQ)	T-Wert
2.1 Textaufbau	1	42%	0.39	1.48	31.1
	2	58%			
2.2 logischer Zusammenhang	1	7%	0.59	1.00	-0.1
	2	52%			
	3	41%			
2.3 sprachlicher Zusammenhang	1	8%	0.64	0.91	-6.3
	2	59%			
	3	33%			

In Tabelle 7 ist die Qualität der Beurteilungskriterien der Dimension «Sprachrichtigkeit» dargestellt. Die Trennschärfe der vier Beurteilungskriterien (Rechtschreibung, Grammatik, Satzbau und Satzzeichen) kann als gut beurteilt werden. Beim Kriterium «Satzzeichen» verteilen sich die Texte relativ gleichmässig auf alle vier Abstufungen, während bei der Rechtschreibung, der Grammatik und beim Satzbau die unterste Stufe (1) eher seltener vertreten ist. Dieses Ergebnis bedeutet, dass wenige Schülerinnen und Schüler die Rechtschreibung, die Grammatik und den Satzbau nur ansatzweise beherrschen, während das korrekte Setzen von Satzzeichen von vergleichsweise etwas mehr Schülerinnen und Schüler nur ansatzweise beherrscht wird.

Tabelle 7: Qualität der Kriterien zur Dimension «Sprachrichtigkeit»

Beurteilungskriterien	Abstufungen	Schwierigkeit	Trennschärfe	Infit (MNSQ)	T-Wert
3.1 Rechtschreibung	1	8%	0.56	1.64	33.9
	2	27%			
	3	46%			
	4	19%			
3.2 Grammatik	1	5%	0.56	1.56	28.1
	2	15%			
	3	41%			
	4	39%			
3.3 Satzbau	1	7%	0.68	1.12	7.3
	2	25%			
	3	47%			
	4	20%			
3.4 Satzzeichen	1	17%	0.56	1.60	32.7
	2	36%			
	3	35%			
	4	12%			

Die Tabelle 8 beinhaltet die Darstellung der Qualität der Beurteilungskriterien der Dimension Sprachangemessenheit. Die Trennschärfe der vier Kriterien ist als gut bis sehr gut zu bewerten und die Kriterien passen gut zum Rasch-Modell (Infit). Bei den ersten beiden Kriterien zum Satzbau und zur Wortwahl sind die Texte hauptsächlich in der mittleren Stufe (2) eingeordnet, während die unterste Stufe (1) seltener vertreten ist. Dies bedeutet, dass die meisten Schülerinnen und Schüler einen abwechslungsreichen Satzbau verwenden und ihre Wortwahl treffend und angemessen ist. Sprachlich sehr sicher und sehr gewandt (Kriterium 4.3, Stufe 4) sind sehr wenige Texte. Dasselbe gilt auch für das letzte Kriterium zum sprachlichen Wagnis und zur Kreativität und Ästhetik. Dieses Ergebnis liegt vermutlich in der Themenwahl begründet: Es ist im Vergleich zu anderen Textsorten schwieriger, in einem öffentlichen Brief die Kreativität zum Ausdruck zu bringen und ein sprachliches Wagnis einzugehen.

Tabelle 8: Qualität der Kriterien zur Dimension «Sprachangemessenheit, Schreibstil und Kreativität»

Beurteilungskriterien	Abstufungen	Schwierigkeit	Trennschärfe	Infit (MNSQ)	T-Wert
4.1 Satzbau	1	9%	0.61	0.95	-2.8
	2	69%			
	3	22%			
4.2 Wortwahl	1	6%	0.64	0.87	-8.3
	2	69%			
	3	25%			
4.3 Schreibstil	1	13%	0.80	0.73	18.9
	2	47%			
	3	38%			
	4	2%			
4.4 Sprachliches Wagnis – Kreativität und Ästhetik	1	6%	0.64	0.93	-4.2
	2	64%			
	3	28%			
	4	2%			

## 4.2 Eindimensionalität des Beurteilungsrasters

Bei der Reliabilität erreicht das gesamte Beurteilungsraster ein sehr gutes Cronbach-Alpha von 0.87. Dies bedeutet, dass Schülerinnen und Schüler mit einer hohen Schreibkompetenz bei allen Kriterien eine höhere Bewertung erhielten als Schülerinnen und Schüler mit einer geringeren Schreibkompetenz. Die Beurteilungskriterien sind somit konsistent angewendet worden und eignen sich gut, um zuverlässig zwischen guten und weniger guten Texten zu differenzieren, und um die Schreibkompetenz von Schülerinnen und Schülern zu bestimmen.

## 5 Beurteilungszuverlässigkeit

### 5.1 Zusammensetzung des Korrekturteams

Die Texte der Schülerinnen und Schüler wurden von sieben Korrektorinnen und Korrektoren (Ratern) bewertet. Rater A studierte Pädagogik und Deutsche Sprach- und Literaturwissenschaft. Rater B ist ausgebildete Primarlehrperson. Rater C studierte Germanistik, Religionswissenschaft und spanische Literatur. Rater D studierte Germanistik, Romanistik und Philosophie und promovierte in Germanistik. Rater E ist ausgebildete Sekundarlehrperson phil. I mit mehrjähriger Erfahrung. Rater G studierte Deutsche Sprach- und Literaturwissenschaften sowie Theaterwissenschaft und Niederlandistik. Rater F studierte Germanistik sowie Medien- und Kommunikationswissenschaften und arbeitet als freie Journalistin. Die Zuordnung der Texte zu den Ratern erfolgte klassenweise nach dem Zufallsprinzip.

## 5.2 Verfahren der Mehrfachkorrekturen

Es besteht die Gefahr, dass sich bei verschiedenen Ratern über die Zeit hinweg ein unterschiedliches mentales Modell einzelner Beurteilungskriterien entwickelt. Daher ist es wichtig in regelmäßigen Abständen das gemeinsame Verständnis der Beurteilungskriterien zu überprüfen (Hoyt, 2000)<sup>8</sup>. Dies wurde dadurch gewährleistet, indem täglich ca. zehn Texte von allen anwesenden Ratern beurteilt wurden (Mehrfachkorrekturen). Das Ziel war dabei, eine möglichst hohe Übereinstimmung der Beurteilung der Texte zwischen den Ratern zu erreichen. Insgesamt wurden mit diesem Verfahren rund 260 Texte von mehreren Ratern beurteilt. Die Abweichungen in der Beurteilung wurden fortlaufend an die Rater zurückgemeldet und mit ihnen besprochen.

## 5.3 Übereinstimmung der Bewertung

In Tabelle 9 sind die Übereinstimmungen beziehungsweise die Abweichungen zwischen den Ratern bei den Mehrfachkorrekturen dargestellt. Die Prozentzahl der Übereinstimmung kommt über einen Vergleich von allen sieben Ratern untereinander zustande (21 mögliche Kombinationen). Die Spalte «Übereinstimmung» beinhaltet den Anteil der vollständigen Übereinstimmung zweier Rater, die darauf folgenden Spalten enthalten dementsprechend den Anteil an Abweichungen um einen, zwei oder drei Punkte pro Kriterium.

Wie Tabelle 9 zeigt, wurde die höchste prozentuale Übereinstimmung beim Kriterium Textaufbau (2.1) erreicht. Dies bedeutet, dass sich die Rater in 86 Prozent der Fälle einig waren, ob ein Text ungegliedert ist, oder ob die Gliederung in Einleitung, Hauptteil und Schluss ersichtlich ist. Die geringste prozentuale Übereinstimmung liegt bei der Beurteilung der Korrektheit des Satzbaus (Kriterium 3.3) vor. Bei diesem Kriterium waren sich die Rater nur in etwa der Hälfte aller Vergleiche einig. Die angestrebte möglichst hohe Übereinstimmung zwischen den Personen des Korrekturteams wurde insgesamt gesehen erreicht. Bei den beiden Kriterien «Textaufbau» (2.1) und «Adressatenbezug» (1.4) wurde die angestrebte Übereinstimmung sehr gut erreicht. Bei den Kriterien «Satzbau: Korrektheit» (3.3) und «Satzzeichen» (3.4) ist die Übereinstimmung hingegen noch verbesserungsfähig.

---

<sup>8</sup> Hoyt, W. T. (2000). Rater bias in psychological research: When is it a problem and what can we do about it? *Psychological Methods*, 5, 64–86.

Tabelle 9: Übereinstimmung und Abweichungen in den Bewertungen

Beurteilungskriterien	Übereinstimmung	Abweichung 1 Punkt	Abweichung 2 Punkte	Abweichung 3 Punkte
1.1 Ausführungen	65%	34%	1%	–
1.2 Fragen	60%	38%	2%	–
1.3 Aussagekraft / Gehalt	62%	37%	1%	–
1.4 Adressatenbezug	79%	20%	0%	–
1.5 inhaltliches Wagnis / Kreativität	60%	39%	1%	0%
2.1 Textaufbau	86%	14%	–	–
2.2 logischer Zusammen- hang	53%	45%	2%	–
2.3 sprachlicher Zusammen- hang	56%	42%	2%	–
3.1 Rechtschreibung	57%	42%	2%	0%
3.2 Grammatik	56%	41%	3%	0%
3.3 Satzbau: Korrektheit	49%	46%	5%	0%
3.4 Satzzeichen	51%	44%	4%	0%
4.1 Satzbau: Abwechslung	61%	38%	1%	–
4.2 Wortwahl	67%	32%	1%	–
4.3 Schreibstil	61%	38%	1%	0%
4.4 Sprachliches Wagnis – Kreativität und Ästhetik	62%	37%	1%	0%

## 5.4 Strenge der Bewertungen

Mit der Anwendung der Item-Response-Theorie<sup>9</sup> ist es zusätzlich möglich, die Korrektur-Strenge der beurteilenden Personen ins Testmodell einzubeziehen und bei der Berechnung der Ergebnisse zu berücksichtigen. Die Strenge der Rater wird dabei auf einer logarithmischen Skala abgebildet und als «Logit» bezeichnet. Ein positiver Logit entspricht einer hohen Strenge, ein negativer Logit einer geringen Strenge. Wie Tabelle 10 zeigt, variiert die Strenge der Rater zwischen Rater A mit

<sup>9</sup> Die statistischen Begriffe in diesem Kapitel werden im Glossar im Anhang dieses Berichtes erläutert.



einem Logit von -0.283 und Rater F mit einem Logit von 0.263. Rater A hat somit am wenigsten streng korrigiert, während Rater F die Texte der Schülerinnen und Schüler am strengsten beurteilt hat. Die Spannweite der Beurteilung beträgt somit 0.55 Logits. Angesichts der Standardabweichung von 1.21 Logits in der Populationsverteilung entspricht dies einem mittelstarken Unterschied in der Korrektur-Strengung der Rater. Damit wird bei der Berechnung der Ergebnisse für die Schülerinnen und Schüler die Rater-Strengung im Modell mitberücksichtigt.

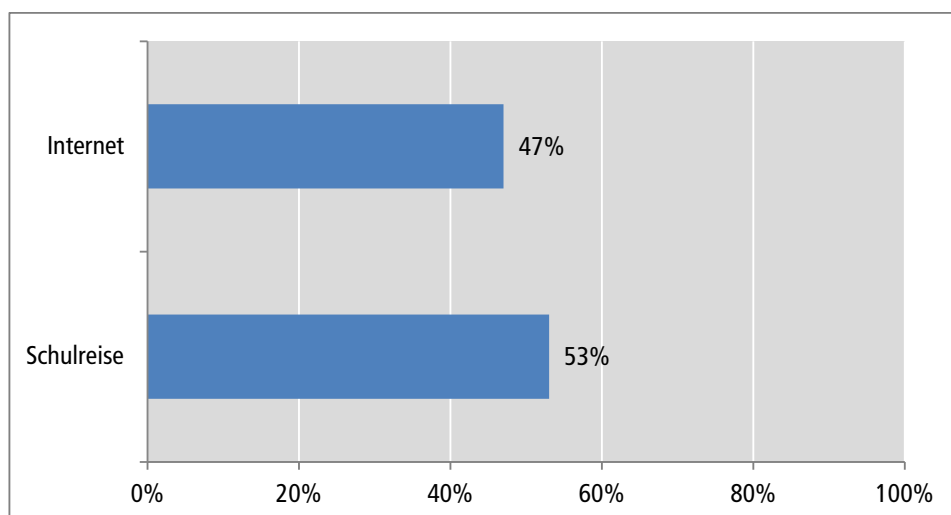
Tabelle 10: Strengung und Modellkonformität der Korrekturen

Rater	Strengung (Logit)	Schätzfehler	Infit (MNSQ)	T-Wert
Rater A	-0.283	0.008	1.00	0.0
Rater B	-0.142	0.008	1.06	1.8
Rater C	0.125	0.008	1.06	1.8
Rater D	-0.103	0.008	1.11	3.1
Rater E	-0.087	0.008	1.06	1.6
Rater F	0.263	0.008	1.07	2.1
Rater G	0.227	0.019	1.08	2.4

## 5.5 Auswahl und Schwierigkeit der Themen

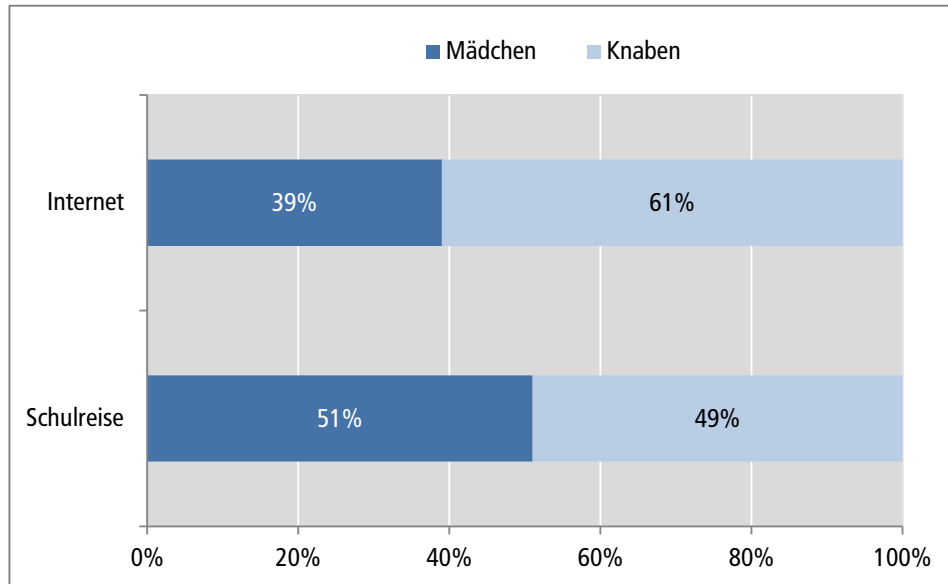
Abbildung 4 zeigt, welches Thema von wie vielen Schülerinnen und Schülern gewählt wurde. Das Thema «Internet» wurde von 47 Prozent der Schülerinnen und Schüler gewählt, 53 Prozent wählten das Thema «Schulreise».

Abbildung 4: Auswahl der Themen



In Abbildung 5 ist die Themenwahl nach Geschlecht dargestellt. Das Thema «Internet» wurde etwas öfter von Knaben gewählt als von Mädchen. Das Thema «Schulreise» wurde in etwa gleich oft von Mädchen oder von Knaben gewählt.

Abbildung 5: Auswahl der Themen nach Geschlecht



Mit Hilfe der Raschskalierung ist es möglich, die Schwierigkeit der beiden Themen zu schätzen und auf einer logarithmischen Skala als «Logit» abzubilden – analog zur Berechnung der Strenge der Rater. Je positiver der Logit ausfällt, desto höher ist die Schwierigkeit eines Themas ausgeprägt (vgl. Tabelle 11). Die beiden Themen unterscheiden sich in der Schwierigkeit nur gering. Es kann nicht ausgeschlossen werden, dass das Thema «Schulreisen» nur deshalb minim schwieriger ausgefallen ist, weil es in der Tendenz eher von schwächeren Schülerinnen und Schülern gewählt wurde. Die Wahl des Themas wird deshalb in den weiteren Analysen beziehungsweise bei der Berechnung der Ergebnisse der Schülerinnen und Schüler nicht berücksichtigt.

Tabelle 11: Schwierigkeit und Modellkonformität der beiden Themen

Thema	Schwierigkeit (Logit)	Schätzfehler	MNSQ	T-Wert
Internet	0.143	0.004	1.01	0.2
Schulreise	-0.143	0.004	1.01	0.4

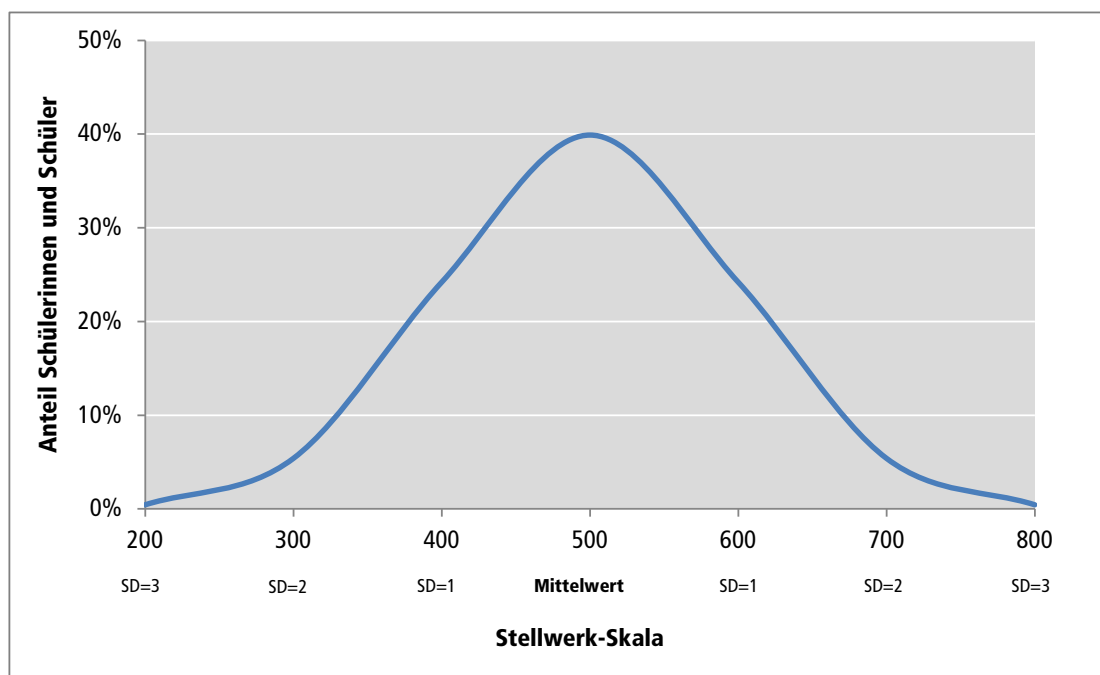
## 6 Testergebnisse

Dass derselbe Text trotz vorgegebener, standardisierter Beurteilungskriterien, sorgfältiger Schulung der Rater und regelmässiger Mehrfachkorrekturen von mehreren Personen immer gleich beurteilt wird, ist aufgrund des Interpretationsspielraums bei offen gestellten Aufgaben nicht zu erwarten. Bei der Beurteilung eines Textes bestimmen vier Faktoren das Testergebnis: a) Die Schreibkompetenz der Schülerin oder des Schülers, b) die Schwierigkeit der Beurteilungskriterien, c) die Korrektur-Strenge der Rater und d) das Thema. Die Fähigkeiten der Schülerinnen und Schüler sowie die Korrektur-Strenge der Rater werden bei der Berechnung der Ergebnisse berücksichtigt.

### 6.1 Verteilung der Ergebnisse auf der Stellwerkskala

Um die Ergebnisse innerhalb des Moduls «Texte schreiben» vergleichbar zu machen, wurden die Fähigkeiten der Schülerinnen und Schüler in eine standardisierte Normalverteilung transformiert, welche analog zur Stellwerk-Skala, einen Mittelwert von 500 und eine Standardabweichung (SD) von 100 Punkten aufweist (siehe Abbildung 6). Diese Skala hat die Eigenschaft, dass rund 68 Prozent der Ergebnisse zwischen 400 und 600 Punkten liegen, rund 95 Prozent zwischen 300 und 700 Punkten und nahezu alle Ergebnisse zwischen 200 und 800 Punkten. Die Punktzahl auf der Stellwerk-Skala zeigt den Schülerinnen und Schülern, wie gut sie innerhalb der Vergleichsgruppe der 10'410 Schülerinnen und Schüler aus der 8. Klasse im Kanton Zürich abschnitten.

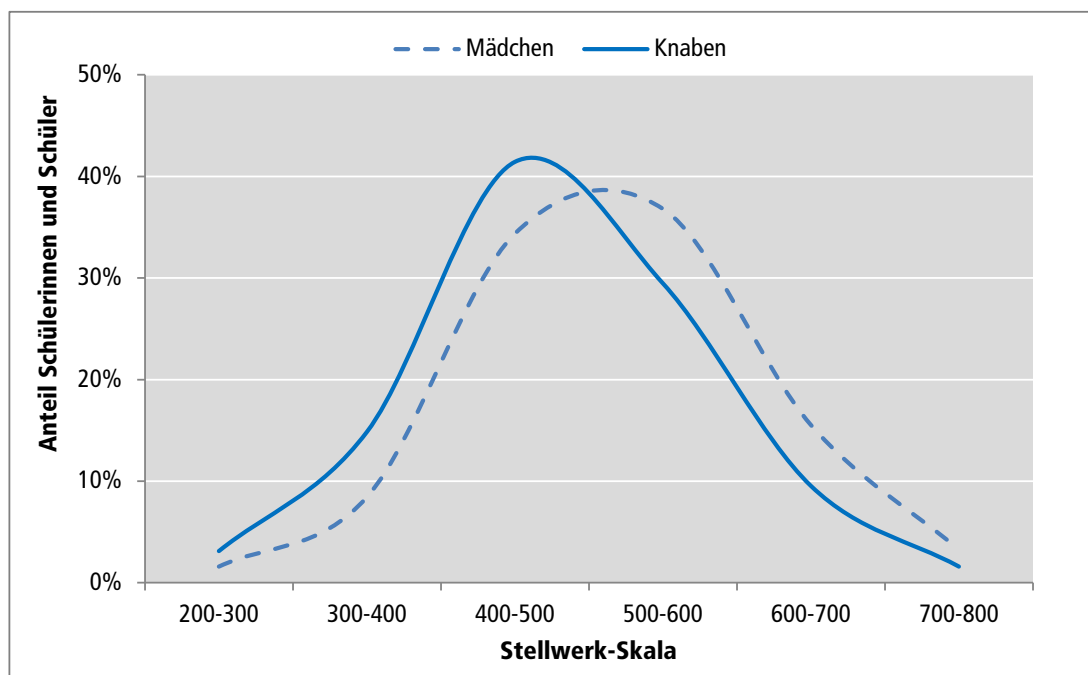
Abbildung 6: Verteilung der Punkte auf der Stellwerkskala



## 6.2 Verteilung der Fähigkeiten der Schülerinnen und Schüler nach Geschlecht

Abbildung 7 zeigt die Verteilung der Fähigkeiten der Schülerinnen und Schüler im Modul «Texte schreiben» als Punktzahl auf der Stellwerkskala, getrennt nach Geschlecht. Die Mädchen erreichen im Durchschnitt 519 Punkte auf der Stellwerkskala, während die durchschnittliche Punktzahl der Knaben bei 485 liegt. Diese Differenz in der Fähigkeit von 34 Punkten ist statistisch signifikant<sup>10</sup> ( $p < .001$ ), jedoch von geringer Bedeutsamkeit (Effektstärke  $d = .35$ ).

Abbildung 7<sup>11</sup>: Fähigkeiten nach Geschlecht im Modul «Texte schreiben»



## 6.3 Verteilung der Fähigkeiten der Schülerinnen und Schüler nach Schulstufe

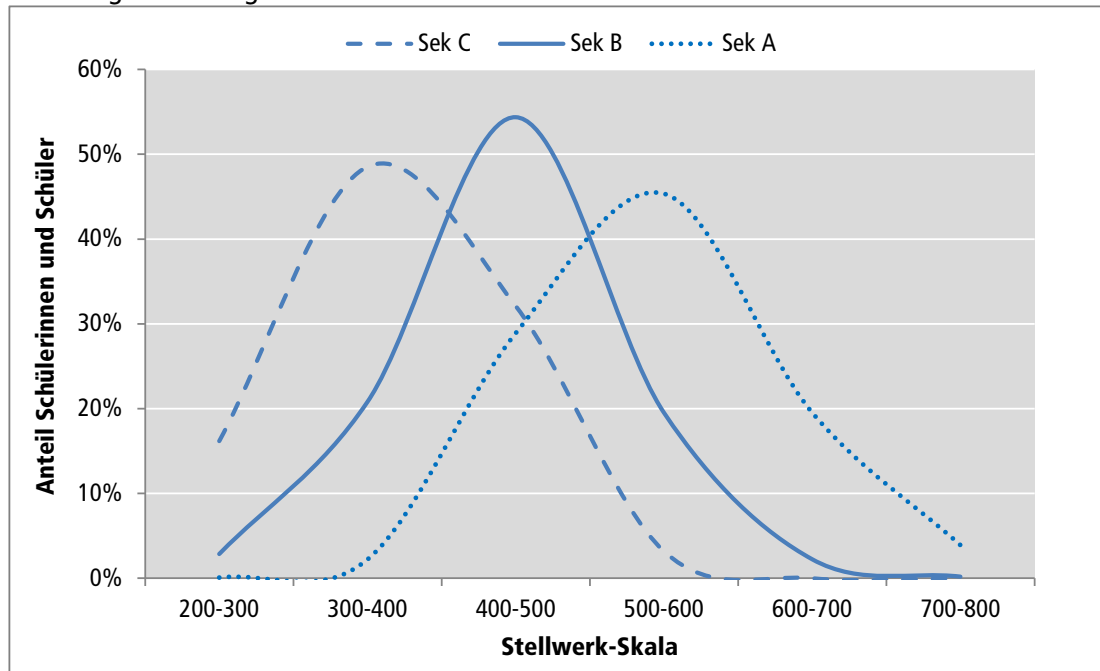
In Abbildung 8 ist die Verteilung der Fähigkeiten der Schülerinnen und Schüler der Sek A, der Sek B und der Sek C im Modul «Texte schreiben» dargestellt.

Wird die durchschnittliche Fähigkeit der Schülerinnen und Schüler getrennt nach Schulstufe betrachtet, so weisen die Schülerinnen und Schüler der Sek C im Durchschnitt 372 Punkte auf. Die durchschnittliche Fähigkeit der Sek B Schülerinnen und Schüler liegt bei 450 Punkten. Diese Differenz von 78 Punkten ist statistisch signifikant ( $p < .001$ ) und sehr bedeutsam (Effektstärke  $d = .99$ ). Die durchschnittliche Fähigkeit der Schülerinnen und Schüler der Sek A liegt bei 547 Punkten. Die Differenz zwischen der Sek A und der Sek B von 97 Punkten ist wiederum statistisch signifikant ( $p < .001$ ) und sehr bedeutsam (Effektstärke  $d = 1.3$ ).

<sup>10</sup> Die statistischen Begriffe in diesem Kapitel werden im Glossar im Anhang dieses Berichtes erläutert.

<sup>11</sup> Anmerkung zur Abbildung 7: Die Darstellung beruht auf einer Kategorisierung der Fähigkeiten der Schülerinnen und Schüler und dient der Illustration der Ergebnisse im Text. Durch die Kategorisierung können die Werte in der Grafik leicht von den exakten Werten im Text abweichen. Dasselbe trifft auch auf die Abbildungen 8 zu.

Abbildung 8: Fähigkeiten nach Schulstufe



Interessant ist, dass 13 Prozent der Schülerinnen und Schüler der Sek C Texte schreiben, deren Beurteilung über dem Mittelwert der Sek B (450 Punkte) liegt. Rund 9 Prozent der Schülerinnen und Schüler der Sek B schreiben Texte, die über dem Mittelwert der Sek A (547 Punkte) liegen. Dieses Ergebnis verdeutlicht, wie sinnvoll eine schultypenunabhängige Beurteilung sein kann.

## 7 Beispieltex te zu den einzelnen Kompetenzniveaus

Um die Kompetenzen der einzelnen Schülerinnen und Schüler standardisiert beschreiben zu können, wurden die Punktzahlen auf der Stellwerkskala zu Kompetenzniveaus zusammengefasst (siehe Tabelle 12). Die Schreibkompetenzen, die eine Schülerin oder ein Schüler innerhalb eines bestimmten Kompetenzniveaus aufweist, wurden zum einen bezüglich der drei Teilgebiete «Inhalt», «Textaufbau» und «Sprache» beschrieben und zum anderen mit Beispieltex ten illustriert. Die Beschreibung der Schreibkompetenzen auf den verschiedenen Niveaus bezieht sich jeweils auf einen durchschnittlichen Text des jeweiligen Niveaus. Bei konkreten Einzelleistungen von Schülerinnen und Schülern kann es vorkommen, dass die Kompetenzen innerhalb einer Gesamtpunktzahl je nach Teilbereich variieren. Beispielsweise könnte ein Text des Intervalls 501 bis 600 Punkte im Teilbereich «Inhalt» tiefer (z.B. im Kompetenzniveau von 401 bis 500 Punkten) und dafür im Teilbereich «Sprache» etwas höher (z.B. im Kompetenzniveau von 601 bis 700 Punkten) liegen. Innerhalb eines bestimmten Kompetenzniveaus besteht somit eine recht grosse Bandbreite an Texten, die sich innerhalb der einzelnen Teilbereiche unterscheiden können. In der folgenden Zusammenstellung wird jedes Kompetenzniveau mit einem Beispieltex t veranschaulicht. Weitere Textbeispiele befinden sich unter [www.ibe.uzh.ch/projekte/texteschreibenzuerich13.html](http://www.ibe.uzh.ch/projekte/texteschreibenzuerich13.html) oder im Praxisteil des vorliegenden Berichts.

Tabelle 12: Kompetenzbeschreibungen nach Punkteintervallen

Punkteintervall	Kompetenzbeschreibungen		
	Inhalt	Textaufbau	Sprache
200 bis 300 Punkte	Die Texte gehen teilweise auf die Aufgabenstellung ein und sind zum Teil gehaltvoll. Teilweise ist auch der Adressatenbezug gelungen. Inhaltlich wird recht viel gewagt.	Der Textaufbau ist zufällig und ungegliedert. Die Sätze und Abschnitte sind teilweise logisch und sinnvoll miteinander verbunden.	Die Rechtschreibung, die Grammatik, die Zeichensetzung und der Satzbau werden ansatzweise beherrscht. Der Satzbau ist etwas abwechslungsreich. Die Wortwahl ist eher treffend und angemessen. Die sprachliche Ausdrucksweise ist recht sicher und recht kreativ.
301 bis 400 Punkte	Die Texte gehen weitgehend auf die Aufgabenstellung ein und sind weitgehend gehaltvoll. Der Adressatenbezug gelingt teilweise. Inhaltlich sind die Texte recht kreativ.	Die Texte sind eher zufällig gegliedert, bestehen aber aus meistens logisch ausgeführten Gedanken. Die Sätze und Abschnitte sind meistens sinnvoll miteinander verbunden.	Die Rechtschreibung, die Grammatik, die Zeichensetzung und der Satzbau werden teilweise beherrscht. Der Satzbau ist recht abwechslungsreich und die Wortwahl ist meistens treffend und angemessen. Sprachlich wird recht viel gewagt. Die Texte sind stilistisch recht sicher.
401 bis 500 Punkte	Die Texte gehen recht gut auf die Aufgabenstellung ein und sind recht gehaltvoll. Der Adressatenbezug ist weitgehend gelungen. Inhaltlich wird viel gewagt.	Die Texte sind gegliedert und bestehen meistens aus logisch ausgeführten Gedanken. Zumeist werden auch die Textelemente sinnvoll miteinander verbunden.	Die Rechtschreibung, Grammatik, die Zeichensetzung und der Satzbau werden weitgehend beherrscht. Der Satzbau ist abwechslungsreich und die Wortwahl treffend und angemessen; der sprachliche Ausdruck ist zumeist sicher, gewandt und recht kreativ.
501 bis 600 Punkte	Die Texte gehen gut auf die Aufgabenstellung ein, sind ausführlich, detailliert, recht gehaltvoll und aussagekräftig. Der Adressatenbezug ist gut gelungen. Inhaltlich sind die Texte kreativ.	In den Texten ist eine Grundgliederung in Einleitung, Hauptteil und Schluss erkennbar. Die Texte bestehen fast immer aus logisch ausgeführten Gedanken und die Textelemente werden fast immer sinnvoll miteinander verbunden.	Die Grammatik und die Setzung der Satzzeichen sind weitgehend fehlerfrei. Die Rechtschreibung wird gut beherrscht. Abwechslungsreich, vielseitig und meistens korrekt ist der Satzbau. Die Wortwahl ist treffsicher und die sprachliche Ausdrucksweise sicher, gewandt und kreativ.
601 bis 700 Punkte	Die Texte gehen ausführlich, detailliert und genau auf die Aufgabenstellung ein, sind gehaltvoll, aussagekräftig und inhaltlich kreativ; unerwartete Ausführungen werden gewagt. Der Adressatenbezug ist sehr gut gelungen.	Die Texte sind in sinnvolle Abschnitte gegliedert, ein roter Faden ist erkennbar. Die Gedankenführung ist logisch und die Textelemente werden auch bei komplexeren Formulierungen sinnvoll miteinander verbunden.	Die Rechtschreibung, die Grammatik, der Satzbau und auch die Setzung der Satzzeichen sind fast fehlerfrei. Die Wortwahl ist auch bei komplexeren Begriffen treffsicher und angemessen. Die Texte enthalten komplexe, abwechslungsreiche Formulierungen, sind sprachlich gewandt und kreativ.
701 bis 800 Punkte	Die Texte gehen sehr ausführlich, detailliert und elaboriert auf die Aufgabenstellung ein, sind sehr aussagekräftig und gehaltvoll. Auch der Adressatenbezug gelingt sehr gut. Inhaltlich enthalten die Texte unerwartete Ausführungen und sind ausgesprochen kreativ.	Die Texte haben einen roten Faden, folgen einem logischen, klaren Aufbau und sind äusserlich und innerlich sinnvoll gegliedert. Die Textelemente werden auch bei komplexeren Verbindungen logisch und sinnvoll miteinander verbunden.	Die Texte sind auch bei komplexeren Formulierungen nahezu fehlerfrei. Die sprachliche Ausdrucksweise ist sehr gewandt und ausdrucksstark. Die Texte zeichnen sich durch ansprechende, abwechslungsreiche und ausgesprochen kreative Sprachstrukturen sowie unerwartete Formulierungen aus.

Textbeispiel 1: 200 bis 300 Punkte

Ich finde, dass das der freie Zugang nicht verboten ist. Der Computer ist auch nützlich in der Schule oder zu Hause. Wenn man bei den Hausaufgaben nicht draus kommt, dann kann man beim Computer nachschauen. In der Schule arbeitet man häufig am Computer, z.B. einen Vortrag machen oder einen Brief schreiben für den Beruf. <sup>das</sup> ~~in~~ ~~den~~ Internet darf man nicht Mössliche Bilder senden oder Schwerttätige ~~Bilder~~ <sup>Videos</sup> schauen dafür ~~darf~~ <sup>kann</sup> man Youtube schauen oder streamen.

Meine Eltern lassen mich an dem Computer um ins Internet zu gehen. Ich bin nicht Computer süchtig, ich lerne an meinem Computer nur für die Schule und auch für die Mofaprüfung. Wenn ich Hausaufgaben bekomme wo ich von der Schule bekomme danach mache ich die Aufgaben die für das Computer zuständig sind. In der Freizeit bleibe ich manchmal zuhause und lerne für die Mofaprüfung oder ich spiele auf einer Seite aber nur etwa 30 Minutenlang. Meine Eltern lassen mich natürlich nicht auf allen Seiten zu gehen, weil sie machen sich sorgen wenn etwas passieren würde. Jeden Tag lerne ich mit dem Computer für die Schule, und nach dem lernen gehe ich mit Freunden Fussballspielen oder wir gehen ins Jugendzentrum [REDACTED]. Ich habe nicht meinen Computer auf dem Händen, ich bin nicht Computer süchtig, für mich ist mein Computer wie die Schule ich lerne drauf. Ich lerne auch mit meine Freunde auch auf dem Computer mehr machen wir nicht.

Freundliche grüsse: [REDACTED]



Liebe Schulleitung

Ich als Schülersprecher, schreibe im Namen der ganzen Klasse und möchte Sie bitten die Abschaffung der Schulreisen und Klassenlager nicht durchzuführen.

Denn man kann bei Schulreisen immer etwas lernen zum Beispiel: Wenn man ins Museum geht oder sonst eine Ausstellung besucht. Ein Klassenlager ist auch gut weil man dort etwas mit der Klasse für eine Woche unternehmen kann und nicht wie sonst immer nur mit dem besten Freund oder mit der besten Freundin.

Um Kosten zu sparen kann man zum Beispiel: mit dem Zug anstatt mit dem Car fahren. Auch eine gute Idee wäre Gruppenangebote wie zum Beispiel: Mit 20 Personen pro Gruppe bekommen sie 20% Rabatt, oder so ähnliche. Für Klassenlager kann man ein günstiges aber gutes Haus mieten, der Nachteil davon ist man muss das Essen selber zubereiten, oder das Lagerhaus gehört der Schule, dann muss man nichts zahlen.

Ich hoffe, ich konnte Sie überzeugen.

Mit freundlichen Grüßen

## Sehr geehrte Eltern

Ich kann euch sehr gut verstehen, wenn sie sich um ihre Kinder sorgen. Das Internet ist sehr gross und dort gibt es viele Seiten, die sie für ihre Kinder vermeiden wollt.

Die Gefahren im Internet sind relativ gross, aber es gibt auch positive Punkte zu diesem Thema. Man kann auch viel lernen, zum Beispiel: könnte man im Internet Vieles über ein Tier, Fahrzeug, Baume etc. erfahren. Man kann sogar ein Fremdwort nachschlagen oder in die eigene Sprache übersetzen. Und wenn man sich einsam fühlt, meldet man sich einfach bei sozialen Netzen an. Es gibt auch viele Menschen, die über das Internet Geld verdienen. Würde ich alles aufzählen, was es kann, müsste ich etwa drei Seiten lang schreiben. Das Internet ist ein Alleskönner!

Was der Sicherheit ihres Kindes im Internet betrifft, gibt es bestimmte Seiten. Diese Seiten sind für solche Fälle spezialisiert, dass man nicht in eine Schuldenfalle hinein gerätet oder sich einen Virus holt. Sonst könnten sie ihr Kind bei Anwesenheit eines erwachsenen Person herum surfen lassen. Natürlich sollte ihr Kind nicht vom Internet umgeben sein. Eins bis zwei Stunden am Tag sind okay. Doch man sollte nicht zu streng sein.

Vielen Dank  
für ihr Verständnis  
euer Schülervertreter

Sehr geehrter Herr [REDACTED]

Wir bedauern es sehr, dass Sie Ausflüge und Klassenlager nicht mehr durchführen wollen.

Wir finden Klassenlager und andere Ausflüge sehr wichtig. Zum Einen, weil man bei solchen Events sein Allgemeinwissen verbessert und zum Anderen, weil der Klassenzusammenhalt gestärkt wird. Wir wissen alle das Sie es nicht mögen, wenn einige ausgeschlossen, ja sogar gemobbt werden. Doch das wird genau passieren, wenn solche Ausflüge gestrichen werden.

Wir verstehen Sie, dass Sie meinen viele wertvolle Lernzeit ginge verloren aber unsere Schule ist immer weit im Lehrplan voraus, deshalb wären ein Klassenlager und einige Ausflüge verkraftbar. Klassenlager sind teuer das wissen alle, jedoch wären unsere Eltern bereit mehr zu zahlen. Die zukünftigen Lager müssen ja nicht mehr in Nobelorten wie z.B. St. Moritz, Davos oder in der Lenzerheide durchgeführt werden. Uns würde auch ein Lager im Tessin oder im Kanton Glarus völlig ausreichen. Auch verstehen wir das Sie die Verantwortung nicht übernehmen wollen. Jedoch würden wir auf actionreiche Events verzichten und so Unfällen vorbeugen.

Wir hoffen sehr, dass Sie Ihre Meinung ändern werden und wir nächstes Jahr ins Klassenlager fahren dürfen.

Mit freundlichen Grüßen

Ihr Schülervertreter der Klassen 2A, 2B und 2C.



Sehr geehrte Schulleitung,  
uns Schülern ist zu Ohren gekommen, dass Sie die in Zukunft stattfindenden Schulreisen abschaffen wollen. Zuerst waren wir sehr bestürzt über ihre Meinung. Doch dann haben wir bemerkt, dass ihre Meinung, aus ihrem Blickwinkel die Beste zu sein scheint. Unsere Stadt muss sparen, das ist uns bewusst, auch das eine Schulreise immer Gefahren mit sich birgt. Auch diese ganze vergeudete Zeit die wir für das Lernen einsetzen sollten. Aber Lernen wir den nichts wenn wir zusammen ein Feuer machen oder im Wald spazieren gehen? Doch, wir lernen für unser Leben. Wir sind knapp bei Kasse? Dann verdienen wir unser Geld. So lernen wir für die Zukunft, dass Geld nicht auf Bäumen wächst. Die Gefahren? Unser Leben besteht aus Gefahren. Es könnte genau in dieser Sekunde in der sich mein Füllfederhalter über das Blatt schwingt, die Decke runterfallen. Sie müssen mir helfen zu verstehen. Niemand von uns versteht. Wieso wird so viel Geld in einen Fussballer investiert und hier haben wir zu wenig Geld für eine Schulreise, geschweige denn ein Lager. Ich habe Glück, ich kann in den Urlaub gehen, Ausflüge machen..

Aber andere nicht, ihnen bleibt, wohl oder übel nur die Schullereise. Jeder Schüler hätte auch ein besseres Bild von der Schule. Nun ist es "das Haus des Lernens und des Spasses" aber wenn die Ausflüge entgültig gestrichen werden sollte, wird es heissen "die Schule der Langeweile, ach ja und des Lernens". Klassenreisen schweissen zusammen. In der Schule bin ich überwiegend ruhig, manchmal muss ich natürlich auch meinen Senk dazu geben, aber nicht so häufig. An den Schullreisen lernen meine Mitschüler meine leicht verrückte Art kennen. Falls es keine Schullreisen mehr geben sollte, würde jeder seine echte Seite zeigen wollen, dann wäre die Schule der reinste Zoo! Wenn Sie einverstanden wären, könnten wir nur eine Schullreise machen, diese einfach genau durchplanen. Wir könnten in unserer Freizeit Kuchen verkaufen und Velo putzen, die Gegend in der wir sein würden in der Schule durchnehmen und so jegliche Gefahren ausschliessen. Im Klassenlager lernt man neue Leute kennen, lernt einander von der Schokolade-Seite kennen, man verliebt sich... Das habe ich alles schon erlebt. Das Klassenlager oder die Schullreise sind Erlebnisse fürs Leben. Hiermit erhoffe ich das ich Sie umgestimmt habe.

Hochachtungsvoll

Stellvertreterin der  
Klasse

## 8 Fazit

Ende November 2013 wurde im Kanton Zürich zum zweiten Mal das Stellwerk-Modul «Texte schreiben» als Ergänzung zu den computerbasierten Stellwerk-Tests flächendeckend durchgeführt. Mit diesem Modul werden die produktiven Fähigkeiten im Fachbereich Deutsch erfasst. Insgesamt nahmen 10'410 Schülerinnen und Schüler der 8. Klasse teil.

Die Schülerinnen und Schüler konnten zwischen zwei verschiedenen Themen wählen: 1) «freier Zugang zum Internet» und 2) «Schulreisen und Klassenlager». Bei beiden Themen wurden die Schülerinnen und Schüler aufgefordert, einen Brief zu verfassen. Drei Fragen dienten den Schülerinnen und Schülern als Leitfaden für das Verfassen der Texte. Ebenfalls zum Schreibauftrag gehörten Angaben darüber, wie die Texte bewertet wurden.

Wie die Analysen zeigten, eignet sich das eingesetzte Beurteilungsraster gut zur Bewertung der Fähigkeiten im Modul «Texte schreiben». Die Beurteilungskriterien sind bis auf wenige Ausnahmen sehr trennscharf. Das heisst, dass die eingesetzten Beurteilungskriterien wesentlich dazu beitragen, die Fähigkeiten der Schülerinnen und Schüler zu bestimmen und zwar so, dass gute Schülerinnen und Schüler eine hohe Punktzahl erreichen und schwache Schülerinnen und Schüler eine tiefe Punktzahl. Die Beurteilungen können aus testtheoretischer Perspektive als zuverlässig bezeichnet werden. Dies bedeutet, dass trotz einer offenen Aufgabenstellung mit einem standardisierten Beurteilungsraster relativ einheitlich beurteilt werden kann.

Die Beurteilungsübereinstimmung der korrigierenden Personen (Rater) wurde täglich überprüft. Die Übereinstimmung zwischen den Ratern im Rahmen der Mehrfachkorrekturen fiel zufriedenstellend aus. Abweichungen im Urteil der Rater liegen zum grössten Teil im Rahmen von plus/minus einem Punkt pro Kriterium. Die bisherige Forschung hat zudem gezeigt, dass es trotz intensiver Schulung und Betreuung der Rater schwierig ist, eine sehr hohe Übereinstimmung zu erzielen (O'Sullivan & Rignall, 2007)<sup>12</sup>. Mit Hilfe der Item-Response-Theorie konnten die leicht unterschiedlichen Beurteilungsmassstäbe der Rater bei der Berechnung der Ergebnisse der Schülerinnen und Schüler korrigiert werden, sodass eine faire Beurteilung gewährleistet ist.

Die Ergebnismeldung auf der Stellwerk-Skala (Mittelwert = 500 Punkte, Standardabweichung = 100 Punkte) darf nicht darüber hinwegtäuschen, dass die Ergebnisse unabhängig von den anderen Testergebnissen der Stellwerk-Tests zu interpretieren sind. Der Mittelwert und die Standardabweichung beziehen sich ausschliesslich auf die 10'410 beteiligten Schülerinnen und Schüler der 8. Klasse im Kanton Zürich. Die Mädchen erreichen im Durchschnitt signifikant bessere Ergebnisse als die Knaben. Die Darstellung der Ergebnisse nach Schultypen zeigt zudem, wie sinnvoll eine schultyp-unabhängige Beurteilung sein kann. Beispielsweise schreiben rund 9 Prozent der Schülerinnen und Schüler der Sek B Texte, welche in ihrer Beurteilung über dem Mittelwert der Sek A liegen.

Abschliessend kann gesagt werden, dass die Erfassung der produktiven Kompetenzen im Fachbereich Deutsch aus einer testtheoretischen Perspektive als zuverlässig betrachtet werden darf. Wenn standardisierte Beurteilungskriterien sowie eine ausführliche Einarbeitungsphase der Raterinnen und Rater mit einer klaren Definition des gemeinsamen Beurteilungsmassstabs eingehalten werden, kann die Erfassung von produktiven Kompetenzen mit der Erfassung von reproduktiven Kompetenzen mittels Leistungstests mithalten.

---

<sup>12</sup> O'Sullivan, B. & Rignall, M. (2007). Assessing the value of bias analysis feedback to raters for the IELTS Writing Modul. In L. Taylor & P. Falvey (Eds.), *IELTS collected papers: Research in speaking and writing assessment* (pp. 446-478). Cambridge, UK: Cambridge University Press.

## 9 Glossar: Statistische Kennwerte

*Effektgrösse* – Punkte-Differenzen werden auf statistische Signifikanz geprüft. Die Signifikanz eines Ergebnisses sagt jedoch nichts aus über dessen Relevanz. Zur Interpretation von statistisch signifikanten Unterschieden wird die Effektgrösse «d» berechnet, indem die Differenz der Mittelwerte durch die Standardabweichungen dividiert wird. Punkte-Differenzen werden so standardisiert und vergleichbar. Eine Effektgrösse von  $d = 0.2$  weist auf einen schwachen Effekt hin, eine Effektgrösse von  $d = 0.5$  auf einen mittleren Effekt und eine Effektgrösse von  $d = 0.8$  auf einen starken Effekt. Eine Punktedifferenz mit einem schwachen Effekt ist für die Praxis wenig bedeutsam, während Punkteunterschiede mit einem mittleren oder starken Effekt als bedeutsam interpretiert werden dürfen.

*Infit (Weighted Mean Square, MNSQ)* – Der Infit (Weighted Mean Square, MNSQ) zeigt, wie gut die einzelnen Beurteilungskriterien zum Rasch-Modell passen. Er zeigt für jedes Kriterium, wie viele unerwartete Bewertungen unter der Annahme des Rasch-Modells beobachtet wurden. Der Infit hat ein Erwartungswert von 1.0. Weichen die Infit-Werte statistisch signifikant von 1.0 ab, dann passt das Beurteilungskriterium nicht zum Rasch-Modell. Ein zu hoher Infit-Wert weist darauf hin, dass die Trennschärfe des Beurteilungskriteriums zu niedrig ist. Ein guter Infit-Wert sollte in der Regel nicht kleiner als 0.7 und nicht grösser als 1.3 sein (Wright & Linacre, 1994)<sup>13</sup>. Der T-Wert zeigt, ob die Infit-Werte statistisch signifikant vom erwarteten Wert von 1.0 abweichen. Ein Wert grösser als 1.96 beziehungsweise kleiner als -1.96 weist auf eine statistisch signifikante Abweichung hin. Die Bewertung eines Beurteilungskriteriums sollte sich allerdings nicht in erster Linie nach dem T-Wert richten, denn dieser hängt unter anderem auch von der Stichprobengrösse ab, welche beim vorliegenden Modul «Texte schreiben» mit 10'410 Schülerinnen und Schüler sehr hoch ausfällt.

*Rasch-Modell* – Die Ergebnisse der Schülerinnen und Schüler werden mit Hilfe der Item-Response-Theorie berechnet. Mit der Anwendung dieses statistischen Modells (Rasch-Modell) ist es zusätzlich möglich, die Korrektur-Strengung der beurteilenden Personen sowie die Schwierigkeit der Themen ins Testmodell einzubeziehen und bei der Berechnung der Ergebnisse zu berücksichtigen. Ein solches Vorgehen wird auch als «Multi-Facetten-Modell» bezeichnet<sup>14</sup>.

*Reliabilität* – Für die Reliabilität beziehungsweise die Messgenauigkeit eines Tests wird der Koeffizient «Cronbach-Alpha» berechnet. Das Cronbach-Alpha zeigt, wie stark die Beurteilungskriterien zusammenhängen. Werte über 0.70 werden als ausreichend, Werte über 0.80 als gut bezeichnet.

*Schwierigkeit* – Die Schwierigkeit eines Beurteilungskriteriums kann durch die Lösungshäufigkeit beziehungsweise den Anteil an Schülerinnen und Schülern, welche die einzelnen Abstufungen erreicht haben, dargestellt werden. Eine Lösungshäufigkeit zwischen 75 und 100 Prozent deutet auf ein (sehr) einfaches Kriterium hin, eine Lösungshäufigkeit von 50 bis 75 Prozent auf ein eher einfaches Kriterium, eine Lösungshäufigkeit zwischen 25 und 50 Prozent auf ein eher schwieriges Kriterium und eine Lösungshäufigkeit von weniger als 25 Prozent auf ein (sehr) schwieriges Kriterium.

*Statistische Signifikanz* – Punkte-Differenzen werden auf statistische Signifikanz geprüft. Eine Differenz ist dann statistisch signifikant, wenn sie durch ein statistisches Testverfahren überprüft

---

<sup>13</sup> Wright, B. D. & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8, 370.

<sup>14</sup> Linacre, J. M. 1994. *Many-Facet Rasch Measurement*. Chicago, IL: MESA Press (original work published in 1989)

und, unter einer im Voraus festgelegten Irrtumswahrscheinlichkeit, beispielsweise 0.1%, für gültig befunden wird.

*Trennschärfe* – Unter der Trennschärfe wird die Korrelation eines Beurteilungskriteriums mit der Gesamtpunktzahl im Modul «Texte schreiben» verstanden. Der Trennschärfekoeffizient zeigt bei einem Test, inwiefern ein Bewertungskriterium Schülerinnen und Schüler mit hoher Punktzahl von Schülerinnen und Schülern mit niedriger Punktzahl trennt. Ein hoher Trennschärfekoeffizient (0.30 bis 1.00) bedeutet, dass Schülerinnen und Schüler mit einer hohen Punktzahl eine hohe Bewertung des jeweiligen Beurteilungskriteriums erreichten und solche mit einer niedrigen Punktzahl eine niedrige Bewertung. Ein niedriger Trennschärfekoeffizient (um 0) bedeutet, dass Schülerinnen und Schüler mit hohen und niedrigen Punktzahlen bei einem Kriterium gleich häufig eine hohe oder niedrige Bewertung erhielten. Ein negativer Trennschärfekoeffizient bedeutet, dass Schülerinnen und Schüler mit einer hohen Punktzahl eine niedrige Bewertung erhielten und solche mit einer niedrigen Punktzahl eine hohe Bewertung. Dementsprechend sollte der Trennschärfekoeffizient nicht kleiner als 0.30 sein. Angewendet auf die Beurteilung von Texten zeigt die Trennschärfe, wie gut die Punktzahl bei einem Kriterium mit der Gesamtpunktzahl übereinstimmt.